– Gedcom –

Service Programs

User Guide

Part 7

- Gedcom Start Center – in Part 1
- Gedcom Conversion – in Part 2
- Gedcom Reduction – in Part 3
- **Gedcom Sorting** V: 4.1.5
- **Gedcom Structure Analysis** V: 4.1.9
- **Gedcom Renumbering** V: 3.1.5
- Gedcom Validation – in Part 4
- Gedcom Duplicate / Merge / Sync – in Part 5
- Gedcom Listings – in Part 6
- GSP – Gedcom Profile – in Part 1

Diedrich Hesmer
Haldenstr 24
71083 Herrenberg
ofb.hesmer@gmx.de

All changes to the previous version are colored

Table of Contents

1   Gedcom Sorting

1.1   Description

Gedcom files (ged-files) are not always sorted according to the respective requirements. To improve this situation, this program has been created.

During sorting the original file will remain unchanged. The name of the new file will be created from the original filename plus "_sor" added and all changed to lower case. So "MyData.ged" will become "mydata_sor.ged". The modified file has the same coding (ANSEL, ANSI or UTF-8) than the original file.

Starting with GEDCOM 7.0 only UTF-8 encoding is allowed, preferably with BOM.

1.2   Files

1.2.1   Sorting Definition file (gssd-file)

This file stores the options and definitions specified at the main screen. They will be created by this program and should be stored using an expressive name. It is recommended to store all gssd-files in a common folder (e.g.. "C:\Data\Gedcom\files\").

1.2.2   Protocol- / Log file

The name of the log file will be composed from the name of the ged-file with an attached "_sor.log", e.g. "my.ged" results in "my.ged_sor.log".

1.2.3   Processing of gdz files

If a gdz file is processed instead of a ged file, after completion of the processing of the ged file contained in the old gdz file a new gdz file is created with the name of the old gdz file, extended by "_sor". This is saved in the folder of the old gdz file. The new ged file created during processing is saved twice

- with the filename of the gdz file with appended "_sor.ged" in the folder of the gdz file. This provides it for the "Edit" button.
- with the same filename of the ged file of the old gdz file in the new gdz file. As standard this is "gedcom.ged".

General information about gdz files and their loading can be found in the User Guide Part 1 in Chapter 5 "Files".

1.3   Checking of DATE values

All DATE lines of the file will be checked. See also GSP User Guide 1 Chapter "Checking of DATE values" and at appendix "Date Formats".

1.4   Main screen

After start of the program the following screen shows up. The name of the last loaded file, if available, and the options used will be reloaded.

- Pushing the respective 📂 a new ged-file or gssd-file will be loaded.
  Note: Each change of the text within the text field for the ged-file, the ged-file will be read to get the tags of the file. During reading several buttons will be disabled. This also will take place when a new gssd-file is loaded.
- Pushing *"Edit"* by the left mouse button opens the ged-file by the text editor for viewing, defined by the "Base-Settings" of the OFB resp. Gedcom Start Center. Alternatively the standard text editor will be used. Pushing the right mouse button starts "*GedShow*" with the ged-file.
- With the optional "*H*" button to the right of the loaded ged file the HEAD record of the ged file is displayed in the browser.
- Select a ***Sorting method*** according the description above.

- Pushing 💾 a folder (e.g. "C:\Data\Gedcom\files\") and name for the Conversion definition file will be defined and the file stored as "gssd-file". This file contains all settings specified by all the tabs for reuse.
- Pushing 🗋 all fields will be set to their initial value.
- A click on the "DropDown" Element for the gssd-selection file opens a list with the recent used gssd files, up to the number defined at "Settings" in the Start-Center. Selecting one of these files will load this file.
- Pushing *?* opens the Gedcom Service Program user guide.
- Pushing the *house* opens the GSP homepage by the browser.
- Pushing *Sort* the reading and sorting of the source file and the writing of the sorted ged-file starts. After completion the new file is sorted in ascending sequence.
- Pushing *Edit log* opens the log-file by the text editor for viewing, defined by the "Base-Settings" of the OFB program resp. Gedcom Start Center. Alternatively the standard text editor will be used.
- Pushing *Close* the program exits and creates an ini-file containing the gssd-filenames for reload.
- Messages will be written to the bottom of the screen.

1.4.1 Tab "Sorting Method + Options"
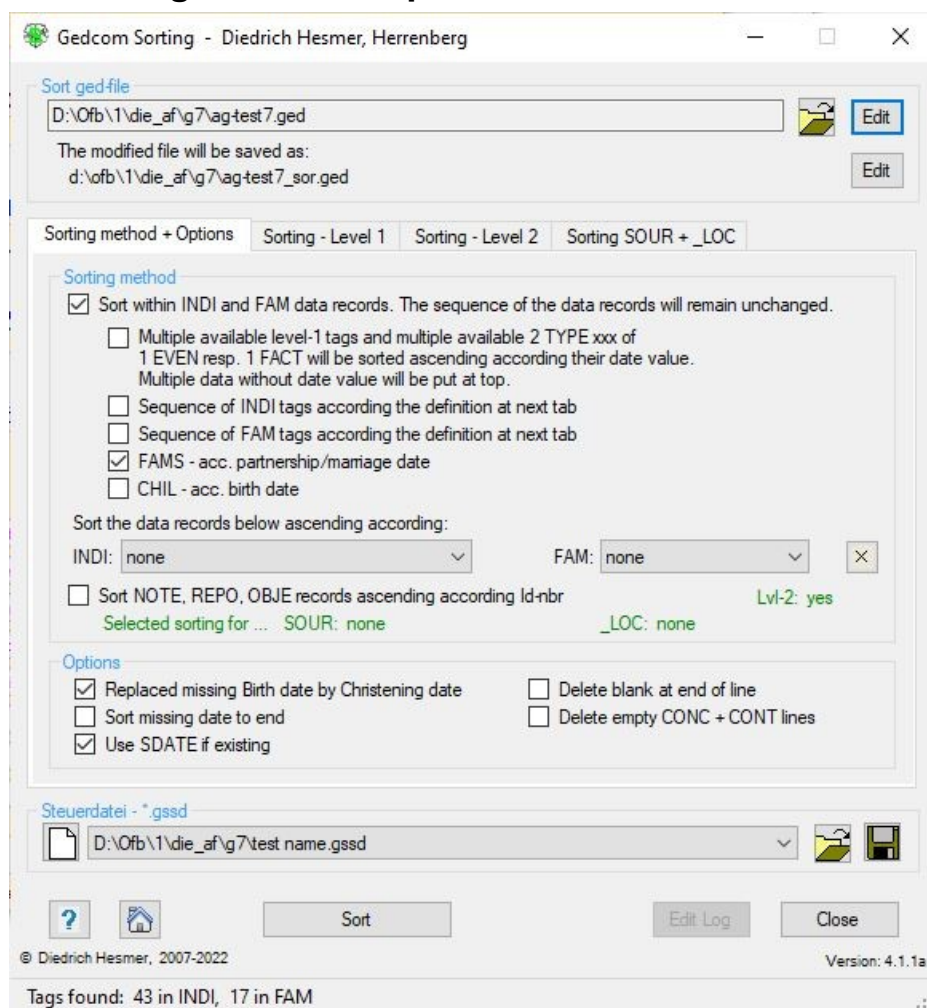
Fig 1.1: Main screen Gedcom Sorting

For a resorting of ged-files. At *Sorting method* following options are available:

- *Sorting within INDI and FAM data records*. The sequence of the data records remain unchanged.
  At least 1 of the sub-options must be selected.
  - ○ *Multiple available level-1 tags* (e.g. "1 OCCU", "1 RESI") and for the tags "1 EVEN" and "1 FACT" multiple available sub-tags "2 TYPE xxxx" (e.g. "1 EVEN" followed by "2 TYPE Education") of the

same data record will be sorted according their DATE tag ascending and new written to the storage device. In case, the DATE tag is missing, the affected event will be moved to the beginning of the affected event group. By this, identical events (e.g. occupation) are in chronological order. The OFB program is processing the ged-data according to their sequence within the ged-file. All other data records remain unchanged.

Not sorted are the tags ADDR AFN ALIA ANCI ASSO CHAN CHIL DESI FAMC FAMS HUSB NAME NCHI NOTE OBJE REFN REPO RESN RFN RIN SEX SOUR WIFE.

- o *Sequence of INDI tags according the ...* Sorting of the level-1 tag sequence within every INDI record according the defined sequence specified by the tab "Sorting – level 1". Within the INDI event and attribute tags also the level-2 tags may be sorted – see tab "Sorting – level 2".

- o *Sequence of FAM tags according the ...* Sorting of the tag sequence within every FAM record according the defined sequence specified by the tab "Sorting – level 1". Within the FAM event tags also the level-2 tags may be sorted – see tab "Sorting – level 2".

- o Sorting of *FAMS* rows ascending according the marriage date. The sequence of FAMS rows within an individual will be modified according the date. All other rows in the file will remain unchanged.

- o Sorting of *CHIL* rows ascending according the birth date. The sequence of CHIL rows within a family will be modified according the date. All other rows in the file will remain unchanged.

- ● *Sort data records below ascending according* allows a sorting of INDI and FAM data records.

  - o *INDI:* The sequence of the data records will be altered. Following sorting options are available

    - ■ none sorting
    - ■ Surname – First name – Birth date – INDI Nbr
    - ■ Surname – Birth date – First name – INDI Nbr
    - ■ INDI Nbr

    The sequence within every INDI data record remains unchanged. All other data records also remain unchanged.

  - o *FAM:* The sequence of the data records will be altered. Following sorting options are available

    - ■ none sorting
    - ■ Marriage date – FAM Nbr
    - ■ FAM Nbr

    The sequence within every FAM data record remains unchanged. All other data records also remain unchanged.

    Note: All dates are used **without** their prefixes, with 2 date values (FROM/TO, BET/AND), the 1st value is used.

- ● *Sort NOTE, REPO, OBJE records ascending according Id-number* modifies the sequence of these records according their Id-number.

  The sequence within every data record remains unchanged. All other data records also remain unchanged.

By this sorting 2 different ged-files may be compared easily by using the program "DiffMerge" or "WinMerge" allowing to find duplicates or to simplify merging of persons or to compare an imported ged-file with the exported one to find out differences and potential data loss.

Date values in the "from..to" format "xxxx / xxxx", "xxxx - xxxx", "/ xxxx" and "xxxx /" will also be processed, even if this is not conform to the Gedcom specification.

The program checks for a correct date. Violations will be written to the log file.

Depending of the selected method following *Options* are available

- o *Replace missing Birth date by Christening date*

- o *Sort missing date to end* sorts entries with missing date to the end of the list for the options "FAMS" and "CHIL". Default is at the beginning of the list.

- o *Use SDATE, if existing* uses the SDATE value (sort date) instead of the DATE value, if existing.

- o *Delete blank at end of line* eliminates all trailing blanks added by some Genealogy-Programs. Excluded are all CONC lines and the line immediate above the CONC lines.

    o **Delete empty CONC + CONT lines** will delete CONT and CONC lines without content, except empty CONT lines in front of CONC.

1.4.2 Tab "Sorting – level 1"

This defines the Output sorting for sorting the tag sequence within every INDI and FAM data record. Only for this the tab area is active.

First select the **Data group** and the **Gedcom Tag Group**, then the **sequence** within the data group.

As **Data group** the INDI data and FAM data tags are available.

As **Gedcom Tag group** the following can be selected:

● **Standard Tags** are predefined by the system and listed as follows:

    o INDI tags:
SEX, BIRT, CHR, BAPM, CHRA, CONF, FCOM, DEAT, BURI, CREM, RELI, EDUC, OCCU, TITL, ADDR, RESI, EVEN, FACT, NOTE, SNOTE, SOUR, OBJE, ASSO, ADOP, NATI, NATU, EMIG, IMMI, CENS, REFN, RIN, CAST, DSCR, PROP, PROB, WILL, GRAD, RETI, ALIA, BLES, ORDN, SSN, IDNO, NCHI, NMR, SUBM, ANCI, DESI, BARM, BASM, BAPL, CONL, ENDL, SLGC, INIL, EXID, NO, UID, _UID

    o FAM tags:
ENGA, MARB, MARR, ANUL, DIVF, DIV, RESI, EVEN, FACT, NOTE, SNOTE, SOUR, OBJE, ASSO, CENS, REFN, RIN, ADDR, MARC, MARL, MARS, NCHI, SUBM, SLGS, EXID, NO, UID, _UID

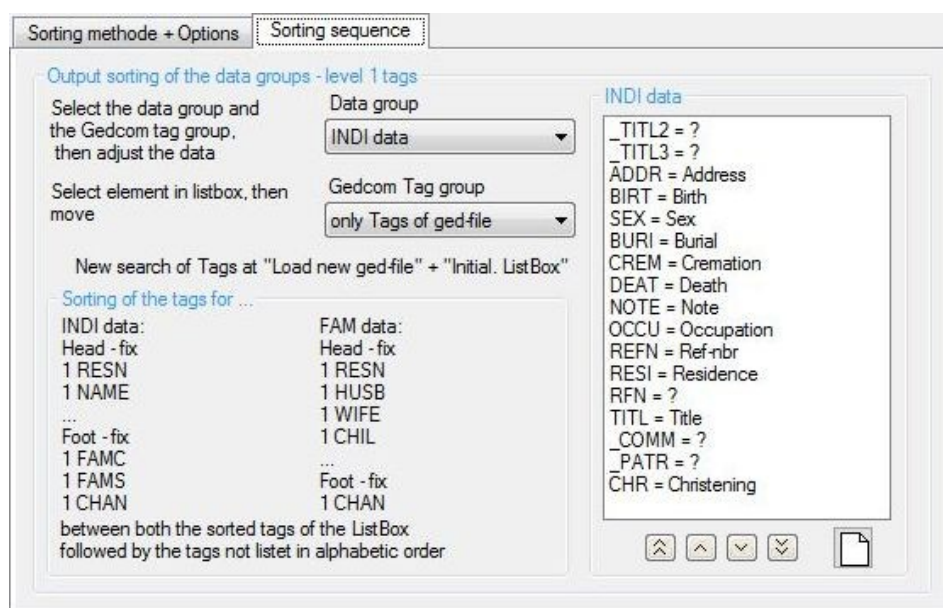   o At the tag ADDR also the tags PHON, EMAIL, EMAI, FAX, WWW, URL will be inserted.

Fig 1.2: Tab "Sorting – level 1"

● **Tags of ged-file** are all the level-1 tags found within the ged-file for INDI and FAM data.

Note: The tags of both groups will be stored in the gssd-file. If during re-use of the gssd-file for the group *Tags of ged-file* additional tags are found in the ged-file, these will be appended to the end of the output list, sorted in alphabetic order.

Note that the sorting of tags will be done by the program as follows:

● INDI data

    o a fix head with
1 RESN, 1 NAME

    o then the tags according the selected sequence will follow.

    o then the not listed level-1 tags including the user defined tags (starting with an underline) will follow.

    o a fix footer with
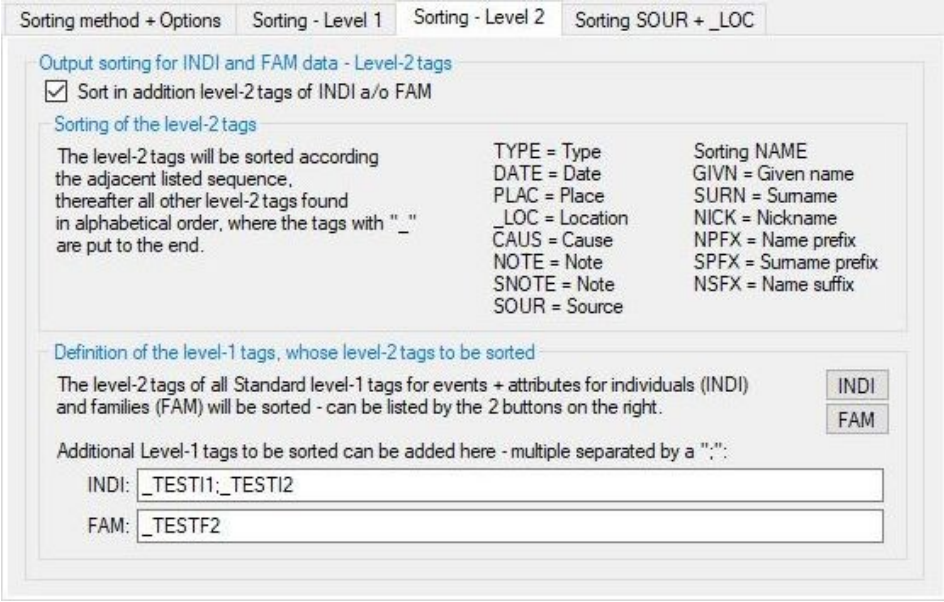
1 FAMC, 1 FAMS, 1 CREA, 1 CHAN
- FAM data
  o a fix head with
    1 RESN, 1 HUSB, 1 WIFE, 1 CHIL
  o then the tags according the selected sequence will follow.
  o then the not listed level-1 tags including the user defined tags (starting with an underline) will follow.
  o a fix footer with
    1 CREA, 1 CHAN

To change the sorting, select a tag in the listbox. The selected element may be moved by ⌃ and ⌄ to the 1st resp last position, by ⌃ and ⌄ just 1 position up resp. down. The sequence within the listbox defines the sequence at the output.

By ☐ the sequence will be set to the system default in case of selected Standard tags. If the Tags of ged-file is selected, the original tags of the ged-file will be listed in alphabetic order.

1.4.3 Tab "Sorting – level 2"

Fig 1.3: Tab "Sorting – level 2"

This allows the definition of the output sorting of level-2 tags for INDI and FAM data records. Only the below described event and attribute tags can be sorted. This tab is only active in case of the proper selected sorting options.

- **Sort in addition level-2 tags ...** activates the fields. This option most likely will be used for comparison of 2 ged-files to adjust the sequence of tags.

- The group box **Sorting of the level-2 tags** defines the tags and their sorting sequence:
  o A fixed sequence od the tags TYPE, DATE, SDATE, PLAC, _LOC, CAUS, NOTE, SNOTE, SOUR. Thereafter all other tags found will be listed in alphabetic order, where tags with a "_" will be put to the end.
  o For the tag NAME the sorting is done in the following order: GIVN, SURN, NICK, NPFX, SPFX, NSFX, then all other tags in the order as found in the ged file.

- The group box **Definition of the level-1 tags, whose level-2 tags to be sorted** explains these tags. By the buttons INDI a/o FAM those level-1 tags, which are sortable according the defined level-2 tags, will be listed by a separate screen. These are only the standard level-1 tags for event and attribute tags.

- o **INDI** level-1 tags:
  ADOP, BAPL, BAPM, BARM, BASM, BIRT, BLES, BURI, CAST, CENS, CHR, CHRA, CONF, CONL, CREM, DEAT, DSCR, EDUC, EMIG, ENDL, EVEN, FACT, FCOM, GRAD, IDNO, IMMI, NATI, NATU, NCHI, NMR, OCCU, ORDN, PROB, PROP, RELI, RESI, RETI, SLGC, SSN, TITL, WILL, INIL

- o **FAM** level-1 tags:
  ANUL, CENS, DIV, DIVF, ENGA, EVEN, FACT, MARB, MARC, MARL, MARR, MARS, RESI, SLGS

- o **Additional level-1 tags to be sorted ...** can be added separately for INDI and FAM. These also will be sorted according the level-2 definitions. Multiple tagnames must be separated by a semicolon (";"). **Not** sortable are the following:

  - ■ INDI level-1 tags:
    ADDR , CHAN, EMAIL, EMAI, FAMC, FAMS, FAX, NAME, PHON, RESN, URL, WWW

  - ■ FAM level-1 tags::
    ADDR, CHAN, CHIL, HUSB, RESN, WIFE

Fig 1.4: Screen "List level-1 tags"

1.4.4 Tab "Sorting SOUR + _LOC"
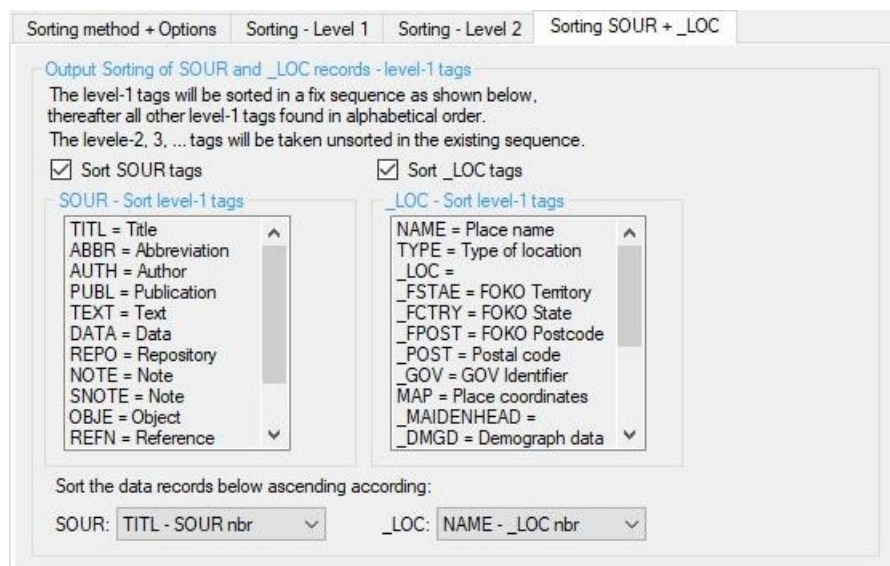
Fig 1.5: Tab "Sorting SOUR + _LOC"

This allows the definition of the output sorting of SOUR and _LOC data records and their level-1 tags. The tags will be sorted in a fixed sequence. These option have been implemented mainly for comparison of 2 ged-files to customize the sequence of records and their tags..

- ● **Sort SOUR Tags** enables the list of tags, which are fixed:
  TITL, ABBR, AUTH, PUBL, TEXT, DATA, REPO, NOTE, SNOTE, OBJE, REFN, RIN, CREA, CHAN

- ● **Sort _LOC Tags** enables the list of tags, which are fixed:
  NAME, TYPE, _LOC, _FSTAE, _FCTRY, _FPOST, _POST, _GOV, MAP, _MAIDENHEAD, _DMGD, _AIDN, NOTE, SNOTE, SOUR, OBJE, EVEN, CREA, CHAN

- ● **Sort data records below ascending according** allows a sorting of SOUR and _LOC data records.

  - o **SOUR:** The sequence of the data records will be altered. Following sorting options are available

    - ■ none sorting

    - ■ TITLE – SOUR nbr

    - ■ SOUR nbr

  The sequence within every SOUR data record remains unchanged. All other data records also remain

unchanged.

- ○ *FAM:* The sequence of the data records will be altered. Following sorting options are available
  - ■ none sorting
  - ■ NAME – _LOC nbr
  - ■ _LOC nbr

  The sequence within every _LOC data record remains unchanged. All other data records also remain unchanged.

2   Gedcom Structure Analysis

2.1   Description

The Gedcom files created by the various Genealogy programs differ in their structure. This program will perform a structure analysis of the ged-file and is intended for use by those users, knowledgeable about ged-files and Gedcom tags to figure out possible errors and specifics of the Gedcom export. During analysis the level-by-level structure of the ged-file for every data record (INDI, FAM, SOUR, NOTE, ...) will be checked.

The analysis allows a marking of conspicuous or faulty tags resp. tag combinations and code values, which do not comply to the 7.x, 5.5.1, the 5.5 or Gedcom-L (group of 23 German speaking program authors) agreements.

During analysis the original file will remain unchanged. The results will be written to screen and to a log-file. The name of the log-file will be created from the original filename plus "_ana.log" added to the end of the filename. So "MyData.ged" will become "MyData_ged_ana.log". This log-file will be added to the folder of the ged-file.

When processing **gdz files**, the contained ged file is extracted and processed accordingly.

General information about gdz files and their loading can be found in the User Guide Part 1 in Chapter 5 "Files".

2.2   Files

2.2.1   Structure Analysis Definition file (gasd-file)

This file stores the options and definitions specified at the main screen. They will be created by this program and should be stored using an expressive name. It is recommended to store all gasd-files in a common folder (e.g.. "C:\Data\Gedcom\files\").

2.2.2   Protocol- / Log file

The name of the log file will be composed from the name of the ged-file with an attached "_ana.log", e.g. "my.ged" results in "my.ged_ana.log".

2.3   Checking of DATE values

All DATE lines of the file will be checked. See also GSP User Guide 1 Chapter "Checking of DATE values" and at appendix "Date Formats".

2.4   Main screen

After start of the program the following screen shows up. The name of the last loaded file, if available at the ini-file, will be reloaded.

● Pushing 📂 a new ged-file resp. gasd-file will be loaded.

● Pushing *"Edit"* by the left mouse button opens the ged-file by the text editor for viewing, defined by the "Base-Settings" of the OFB resp. Gedcom Start Center. Alternatively the standard text editor will be used. Pushing the right mouse button starts "*GedShow*" with the ged-file.

● With the optional "*H*" button to the right of the loaded ged file the HEAD record of the ged file is displayed in the browser.

● Pushing 💾 a folder (e.g. "C:\Data\Gedcom\files\") and name for the Analysis definition file will be defined and the file stored as "gasd-file". This file contains all settings specified by all the tabs for reuse.

● Pushing 🗋 all fields will be set to their initial value.

● A click on the "DropDown" Element for the gasd-selection file opens a list with the recent used gasd files, up to the number defined at "Settings" in the Start-Center. Selecting one of these files will load this file.

● Pushing *?* opens the Gedcom Service Program user guide.

● Pushing the *house* opens the GSP homepage by the browser.

- Pushing *Analyse* starts the structure analysis. The results will be written to screen and log-file.
- Pushing *Edit log* opens the log-file by the text editor for viewing, defined by the "Base-Settings" of the OFB program resp. Gedcom Start Center. Alternatively the standard text editor will be used.
- Pushing *Delete log* erases the stored log-file.
- Pushing *Close* the program exits and creates an ini-file containing the ged-filename and the options for reload.
- Messages will be written to the lower screen part.
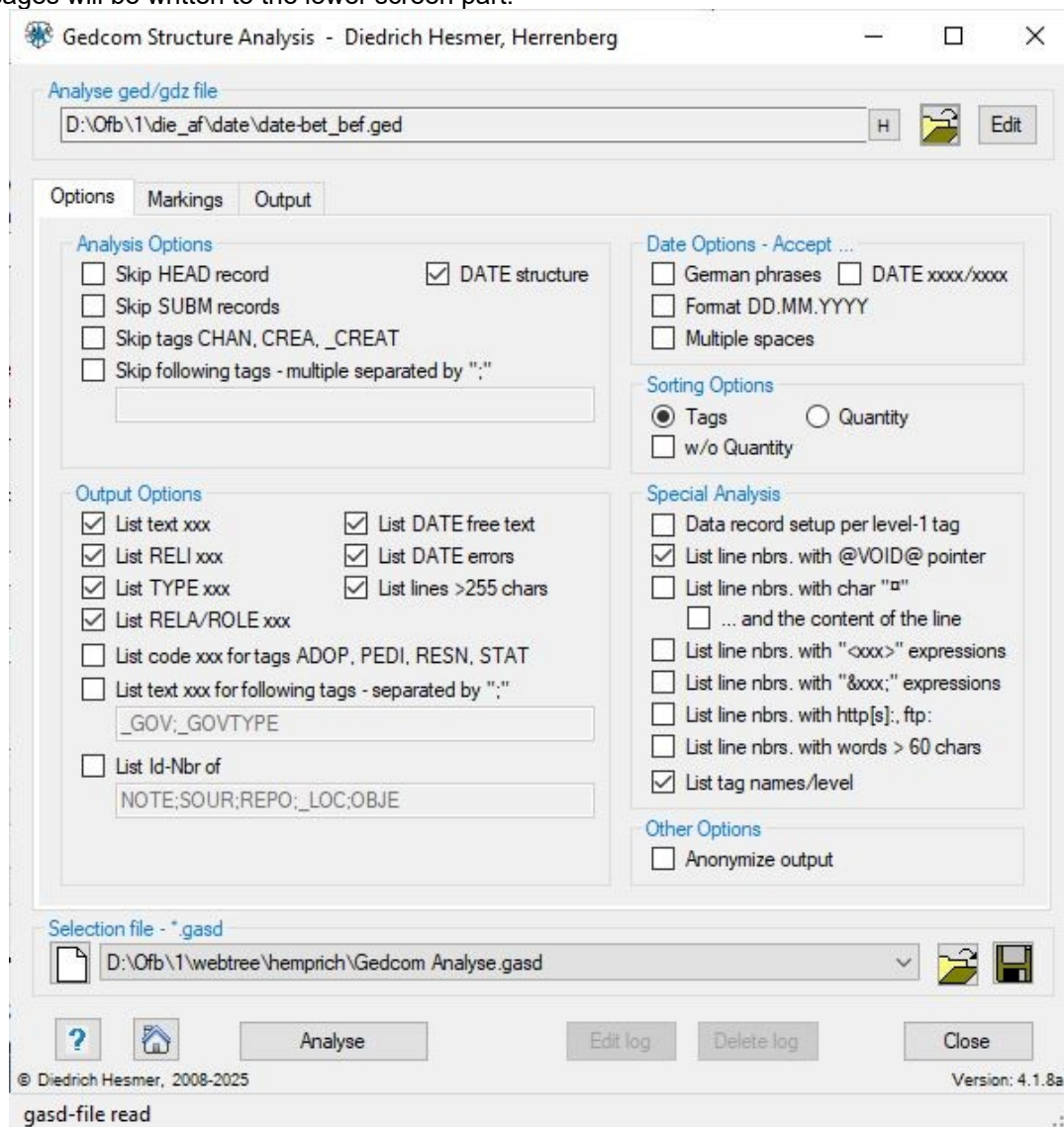
Fig 2.1: Main screen "Gedcom Structure Analysis" + Tab "Options"

2.4.1 Tab "Options"

- The group *Analysis Options* affects the range of the analysis.
    - *Skip HEAD records* skips the analysis of this record.
    - *Skip SUBM records* skips the analysis of these records.
    - *Skip tags CHAN, CREA/_CREAT* skips the analysis of these tags.
    - *Skip following tags – separated by ";"* allows to add tags, which should skipped during analysis. The tags must be CAPITALIZED and separated by ";".
    - *DATE Structure* performs a structure analysis of all DATE lines. They will be checked for conformance to the Gedcom Standard, Erroneous lines optional can be listed and the number of error-

free DATE lines for every DATE prefix (ABT, FROM, BEF, ...) will be listed.
Note: So called free form text, typified by completely bracketed text (example: "2 DATE (Eastern 1920)"), will not be analyzed, but optional listed.

- The group *Output Options* defines the content of the output
  - *List Text xxx* lists those tags with an attribute "text" following in addition to the tags without this data (e.g. "NOTE" and "NOTE text".
  - *List RELI xxx* lists the RELI tags with their attribute "text" separately for every "text" (e.g. "RELI baptism").
  - *List TYPE xxx* lists the TYPE tags with their attribute "text" separately for every "text" (e.g. "TYPE occupation" and "TYPE house").
  - *List RELA/ROLE xxx* lists the REALA resp. ROLE tags with their attribute "text" separately for every "text" (e.g. "RELA godfather" and "ROLE witness").**.**
  - *List code xxx for tags ...* lists the attribute "text" for the tags ADOP, PEDI, RESN + STAT separately for every "text".
  - *List text xxx for following tags – separated by ";"* allows the input of additional tags with coded values. The tags must be CAPITALIZED and separated by ";".
  - *List Id-Nbr. of* allows the output of record id's found. The record tags must be CAPITALIZED and separated by ";".
  - *List FILE path xxx* lists the found path types of FILE tags, separated for each of the following types:
    - **"_:/xxx":** Absolute path reference with a drive letter.
    - **"//xxx":** Network path reference, e.g. local network.
    - **"/xxx":** Absolute path reference.
    - **"./xxx":** Relative path reference, on local PC.
    - **"../xxx":** Relative path reference, on local PC.
    - **"://xxx":** Web path reference, e.g. http://xxx.
    - **"xxx":** all others, no path reference, only file name.
  - By *List DATE free text* all lines will be listed, where the whole text is bracketed and defined by this as "free form text".
  - By *List DATE errors* all lines, not conformance to the Gedcom standard, will be listed.
  - By *List lines >255 chars* all lines with more than 255 characters will be listed az the end of the log-file.
- Group *Date Options – Accept ...* affects the date checking.
  - *German text phrases* rates following texts as correct:
    - Months shortcuts: "MÄR", "MRZ", "MAI", "OKT", "DEZ".
    - Date prefixes: "AB", "BIS", "CA", "CA.", "ERR", "ERR.", "GESCHÄTZT", "NACH", "SEIT", "UM", "VON", "VON xxx BIS", "VOR", "ZW xxx UND", "ZW. xxx UND", "ZWISCHEN xxx UND".
  - *Format DD.MM.YYYY* rates this German notation as correct. As separator only the "." will be accepted.
  - *DATE xxxx/xxxx* accepts date values in the form "year/year", 4 digits each.
  - *Multiple spaces* accepts date values with multiple spaces.
- The group *Sorting Options* defines the sorting of the output list
  - "*Tags"* sorting according tag name.
  - "*Quantity"* sorting according the numbers of tag combinations.
  - *w/o Quantity* suppresses the output of the quantity of tag combinations.
- The group *Special analysis* allows the following checks
  - *Data record per level-1 tag* is based on the standard analysis. For each level-1 tag the sub-structures will be listed appended in 1 line.

    ○ ***List line nbrs. with VOID@ pointer*** will list all line numbers where @VOID@ pointer (placeholder) were found, example "1 CHIL @VOID@" or "2 ASSO @VOID@".

    ○ ***List line nbrs. with char "¤"*** will list all line numbers where the character was found. This is helpful for the OFB and some of the Gedcom Service Programs, this character is used as internal control character and therefor not allowed in the ged-file.

       ■ ***...and the content of the line*** lists in addition to the line number the text of the line.

    ○ ***List line nbrs. with "<xxx>" expressions*** lists all expressions in angle brackets "<xxx>" (e.g. HTML commands "<br>", "</p>", ...), found in NOTE, SOUR and TEXT and their CONC and CONT sub-structures with line numbers. The expressions found in CONC lines will be counted at the NOTE, SOUR, TEXT resp. CONT line.

    ○ ***List line nbrs. with "&xxx;" expressions*** lists all text entities "&xxx;" (e.g. "ä" for "ä"), found in NOTE, SOUR and TEXT and their CONC and CONT sub-structures with line numbers. The expressions found in CONC lines will be counted at the NOTE, SOUR, TEXT resp. CONT line.

    ○ ***List line nbrs. with http[s]:, ftp:*** lists all http:, https: and ftp: expressions, found in NOTE, SOUR and TEXT and their CONC and CONT sub-structures with line numbers. The expressions found in CONC lines will be counted at the NOTE, SOUR, TEXT resp. CONT line.

    ○ ***List line nbrs. with words > 60 chars*** lists all words > 60 characters with line numbers and tags.

    ○ ***List tag names/level*** selects for every level number the tags and lists the tags found, except HEAD, TRLR and i.a. the optional skipped tags/records. Unknown tags not starting with a "_" will be marked by a "#" as 1st character.

● The group ***Other Options*** includes:

    ○ Anonymize output allows an anonymization of output lists by:

       ■ Suppresses for the failure list the output of "[NAME]".

       ■ For lines >255 chars only the first + last 5 chars will be listed.

2.4.2 Tab "Markings"

Fig 2.2: Tab "Markings"

Here the options for checking tags, tag combinations and checking codes and predefined text will be defined.

● The group ***Marking of lines containing … Tags*** describes the marking of output lines of the structure analysis, if affected. Multiple characters may be listed.

    ○ ***Checking according GEDCOM*** shows the detected GEDCOM version of the loaded ged file. This is 7.x for version 7.0 and newer, 5.5.1 for version 5.5.1 and older or ??? as system value.

o ***Marking of special structures according detected GEDCOM*** marks the following structures:

- ■ **"?"** for unknown tags, not complying to the standard.
- ■ *"X"* for questionable tag combinations, if e.g. the tag AGE is found as level-1 tag in a FAM record.
- ■ *"F"* for failure in structure – level number of next line increased by >1.
- ■ *"T"* for missing text after the tag, e.g. the occupation is missing after the tag OCCU.
- ■ *"@"* for a missing reference Id, e.g. if after ALIA no "@Ixx@" is found.
- ■ *"Y"* for faulty text after a tag. After BIRT, CHR and MARR no normal text is allowed, but noting or a single "Y".
- ■ *"N"* for not allowed text after the tag.
- ■ *"-"* for none tested tags or tag combinations.
- ■ **"#"** for the FILE tag, if the path starts with "~\". This indicates to a ged-file created by a Mac program, e.g. "Reunion", and means the "Home" directory of the Mac.

Note: When selecting this option, automatically the option "*List Text xxx*" will be activated.

All essential tags and tag combinations will be checked except:

- ● CHAN structures
- ● Structures of roman (ROMN) and phonetic (PHON) language variations

o ***Include Gedcom 5.5*** checks and marks in addition to 5.5.1 also the tags and codes according 5.5. Not enabled for GEDCOM 7.x

o ***Marking of codes and values according detected GEDCOM*** marks those code resp. value entries, which do NOT comply to Gedcom 5.5.1.

- ■ *"U"* for unknown codes/values. Also those data will be marked, which are accepted as user-defined values.
- ■ ***Code check for 5.5.1 independend of upper/lower case*** accepts valid code values that are completely in lower case or completely in upper case.
- ■ ***Tags no checking:*** The check for non Gedcom-compliant values can be suppressed for the tags listed below.
  - ● *AGE*
  - ● *ROLE*
  - ● *LATI + LONG*

All essential tags containing codes will be checked. Following groups will **not** be checked:

- ● DATE values other than Gregorian
- ● Language codes inclusive roman and phonetic values
- ● Multimedia format and media types
- ● TIME values

o ***Marking of Gedcom-L structures / codes/values*** marks those tags resp. tag combinations and code resp. value entries, where the Gedcom-L group has defined agreements to improve the import and export between the concerned programs. The markings are – all in small letters with a circumflex:

- ■ *"ŝ"* for Gedcom-L conformable structures.
- ■ *"ĉ"* for Gedcom-L conformable codes/values.

All essential tags containing codes will be checked. Following groups will **not** be checked:

- ● DATE values other than Gregorian
- ● Language codes inclusive roman and phonetic values
- ● Multimedia format and media types
- ● TIME values

o ***Marking of potential OFB structures for …*** will support you in creating necessary Profile files and naming files for the OFB program. It checks the tags and tag combinations in reference to possible characteristics for the OFB program. The markings are – all in small letters:

- **"p"** points to a fact, that this tag combination might be defined by the profile file. These are lines, where tags start with "_" or which have a sequence of characters of "ADDR", "ADOP.PLAC", "ALIA @", "ASSO @", "BAPM", "RESI".
- **"t"** marks every line with the TYPE tag. It also points to a fact, that this tag combination might be defined by the profile file and/or the text of the TYPE tag might be modified by the naming file.
- **"c"** marks every line with the tag ADOP, AGE, MEDI, STAT or the tag combination _LOC.NAME.LANG where the text of the tag might be modified by the naming file.
- **"r"** marks every line with the tags ROLE or RELA where the text of the tag might be modified by the naming file.
- **"a"** marks every line instead of "*p*", which might be defined for the profile as "Tags 1/2 added". These are especially lines with unknown tags and tags starting with "_". Alternatively these might be defined as "Special Group".
- **"s"** marks every line instead of "*p*", which might be defined for the profile as "Special Group".
- **"o"** marks every line with none-standard Gedcom tag combinations, but accepted by the OFB.
- **"n"** marks every line with tags not supported by the OFB and therefor not definable by the profile. In case of a demand, please contact me.

This option only is active if above "Marking of special structures" is selected.

- The group ***List markings in screen*** defines the range of the listing of the markings.
  - ***All possible*** lists all possible markings in a separate screen.
  - ***Only found*** lists only the found markings in a separate screen.
  - The button ***Edit Mark*** opens a separate screen Fig 2.4: Screen "Explanation of Markings" with the explanation of the used codes.

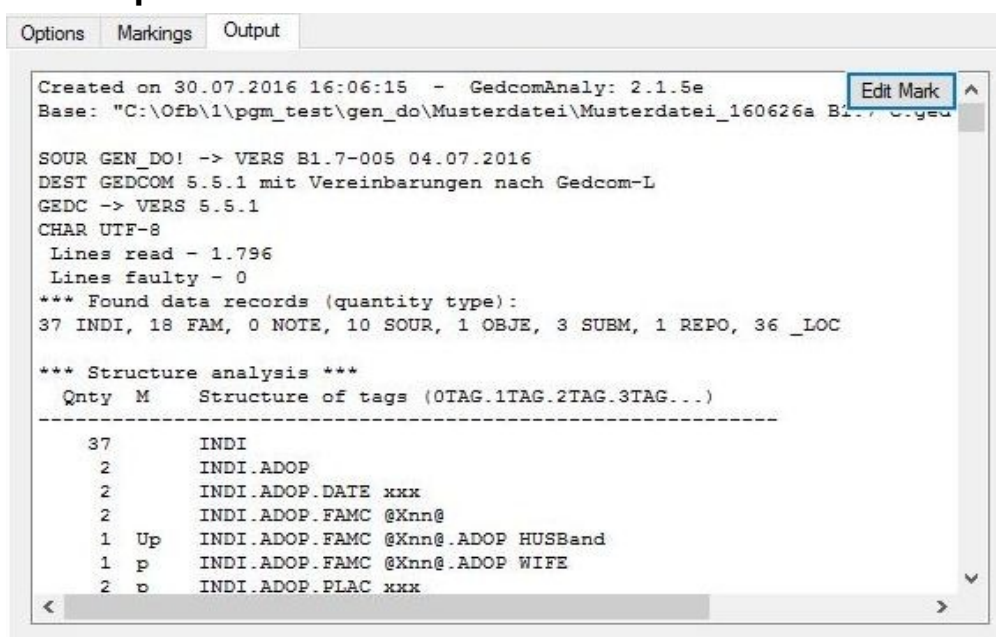## 2.4.3 Tab "Output"

Fig 2.3: Tab "Output"

The output tab contains the analysis results. The same data will be written to the log-file.

The button ***Edit Mark*** opens a separate screen Fig 2.4: Screen "Explanation of Markings" with the explanation of the used codes.

Following information will be listed at the output:

- Date and time of the analysis
- Name and creation time of the ged-file

● Possible errors/faults of the ged-file, including line number and content
● Number of found records per type
● Results of the structure analysis. Each structure and the frequency "n" will be listed and the representation in the list will be for the Structure Analysis

```
Qnty  Structure of tags (0TAG.1TAG.2TAG.3TAG...)
----------------------------------------------------------
n INDI
n INDI.NAME
n INDI.SEX
n INDI.BIRT.DATE
n INDI.BIRT.PLAC
n ...
```

Fig 1: Tag structures

```
Qnty  Structure of tags (0TAG.1TAG.2TAG.xTAG...2TAG.xTAG...)
----------------------------------------------------------
n INDI
n INDI.NAME
n INDI.SEX
n INDI.BIRT.2DATE.2PLAC.2SOUR.2NOTE.3CONT
n INDI.DEAT.2DATE
n INDI.RELI
n INDI.SOUR
n INDI._BURGERORT
n INDI.FAMS
```

Fig 2: Level-1 analysis

```
ged-file representation              results in   structure analysis representation
0 @I15@ INDI                                      n INDI
1 NAME Marion /Afrau/                             n INDI.NAME
1 SEX F                                           n INDI.SEX
1 BIRT                                            n INDI.BIRT
2 DATE 16 MAY 1888                                n INDI.BIRT.DATE
2 PLAC Aldorf                                     n INDI.BIRT.PLAC
2 SOUR Standesamt Hadorf                          n INDI.BIRT.SOUR
2 NOTE Geburtsregister ...                        n INDI.BIRT.NOTE
3 CONT Fortsetzung Text ...                       n INDI.BIRT.NOTE.CONT
1 DEAT                                            n INDI.DEAT
2 DATE 13 NOV 1976                                n INDI.DEAT.DATE
...                                               n INDI.RELI
1 RELI evangelisch                                n INDI.SOUR
1 SOUR #0                                         n INDI._BURGERORT
1 _BUERGERORT Hadorf                              n INDI.FAMS
1 FAMS @F4@
```

Fig 3: Presentation of Structure Analysis

```
Qnty  Structure for all error-free DATE lines
---------------------------------------------------------
    2  ABT xxx
    2  AFT xxx
    2  BEF xxx
    0  BET xxx AND xxx
    0  CAL xxx
    2  EST xxx
    0  FROM xxx
    0  FROM xxx TO xxx
    0  INT xxx
    1  SORT xxx
    0  TO xxx
   33  without prefix xxx
42 error-free DATE lines in file
```

Fig 4: DATE Structure Analysis

```
*** Free text DATE lines
 Line: Content
---------------------------------------------------------
20: @I12@ -> 1 DEAT => 2 DATE (Eastern 1698)
28: @I25@ -> 1 DEAT => 2 DATE (after 1659, before 1670)
42: @I34@ -> 1 DEAT => 2 DATE (at 4.1669)
84: @I38@ -> 1 DEAT => 2 DATE (1738 or 05.1741)
...
13 lines with free text
```

Fig 5: Date free form text listing

```
Qnty  Structure of tags (0TAG.1TAG.2TAG.xTAG...2TAG.xTAG...)
---------------------------------------------------------
n INDI
n INDI.NAME
n INDI.SEX
n INDI.BIRT.2DATE.2PLAC.2SOUR.2NOTE.3CONT
n INDI.DEAT.2DATE
n INDI.RELI
n INDI.SOUR
n INDI._BURGERORT
n INDI.FAMS
```

Fig 6: Level-1 analysis

```
*** Found Tag Names ***
Level
Nr Cnt:  Tag names
---------------------------------------------------------
 0   7:  FAM INDI NOTE OBJE REPO SOUR #LABL
 1  59:  ABBR ADDR ADOP ASSO AUTH BIRT BURI CHIL CHR CONF
         ...
         SUBM TEXT TITL WIFE WILL _CRE #COLR #MRCI #MRRE
         ...
 3  11:  DATE MAP PAGE _INFO _LAD _LAM _LAS _LOD _LOM _LOS
 4   3:  LATI LONG TIME
89 different tags in file - without HEAD TRLR CHAN _CREAT
  --> 74 Standard Tags, 10 _XXX User-defined Tags, 5 #-marked Unknown Tags
```

Fig 7: Found Tag Names

2.4.4 Screen "Explanation of Markings"

This screen lists dynamically the markings of the selected resp. found options. It is a permanent screen and
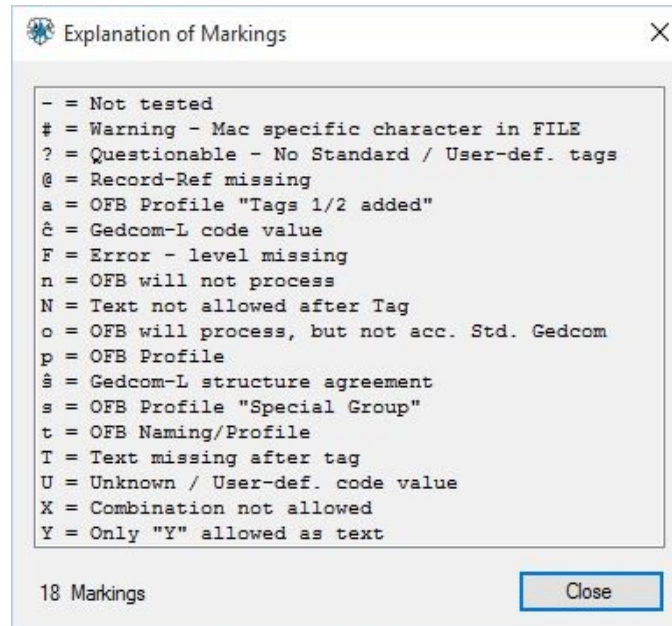

Fig 2.4: Screen "Explanation of Markings"

can be moved to any place of the screen to be visible side by side to the output screen.

The button **Close** closes the screen.

3   Gedcom Renumbering

3.1   Description

Gedcom files (ged-files) are not always continuously numbered according to the respective requirements or are very cryptic. To improve this situation, this program has been created. In addition to the renumbering of below stated records the individual records may be sorted by different ways.

During renumbering the original file will remain unchanged. The name of the new file will be created from the original filename plus "_num" added and all changed to lower case. So "MyData.ged" will become "mydata_num.ged". The modified file has the same coding (ANSEL, ANSI or UTF-8) than the original file.

Starting with GEDCOM 7.0 only UTF-8 encoding is allowed.

Independent of the sequence of records of the original file, the sequence of the modified file will be:

● Header with HEAD, SUBM and SUBN, if existing

● INDI – sorted if required

● FAM

● NOTE resp. SNOTE

● SOUR

● REPO

● _LOC

● OBJE

● Further additional existing record types

● Level-0 tags, which may contain sub-tags, not being a record

● Finally TRLR

The records for FAM, NOTE, SOUR, REPO, _LOC, OBJE and further additional existing are not sorted. The sequence within each group is according to the sequence of the original file, except for sorted INDI records.

@VOID@ Pointers remain unchanged.

3.2   Files

3.2.1   Renumbering Definition file (gnsd-file)

This file stores the options and definitions specified at the main screen. They will be created by this program and should be stored using an expressive name. It is recommended to store all gnsd-files in a common folder (e.g.. "C:\Data\Gedcom\files\").

3.2.2   Protocol- / Log file

The name of the log file will be composed from the name of the ged-file with an attached "_num.log", e.g. "my.ged" results in "my.ged_num.log".

3.2.3   Processing of gdz files

If a gdz file is processed instead of a ged file, after completion of the processing of the ged file contained in the old gdz file a new gdz file is created with the name of the old gdz file, extended by "_num". This is saved in the folder of the old gdz file. The new ged file created during processing is saved twice

● with the filename of the gdz file with appended "_num.ged" in the folder of the gdz file. This provides it for the "Edit" button.

● with the same filename of the ged file of the old gdz file in the new gdz file. As standard this is "gedcom.ged".

General information about gdz files and their loading can be found in the User Guide Part 1 in Chapter 5 "Files".

3.3   Main screen

After start of the program the following screen shows up. The name of the last loaded file, if available, and the options used will be reloaded. At the message line a request will be displayed to move the cursor within the screen. This will initiate a first reading of the ged-file to find out the record types stored. These numbers will be displayed here.
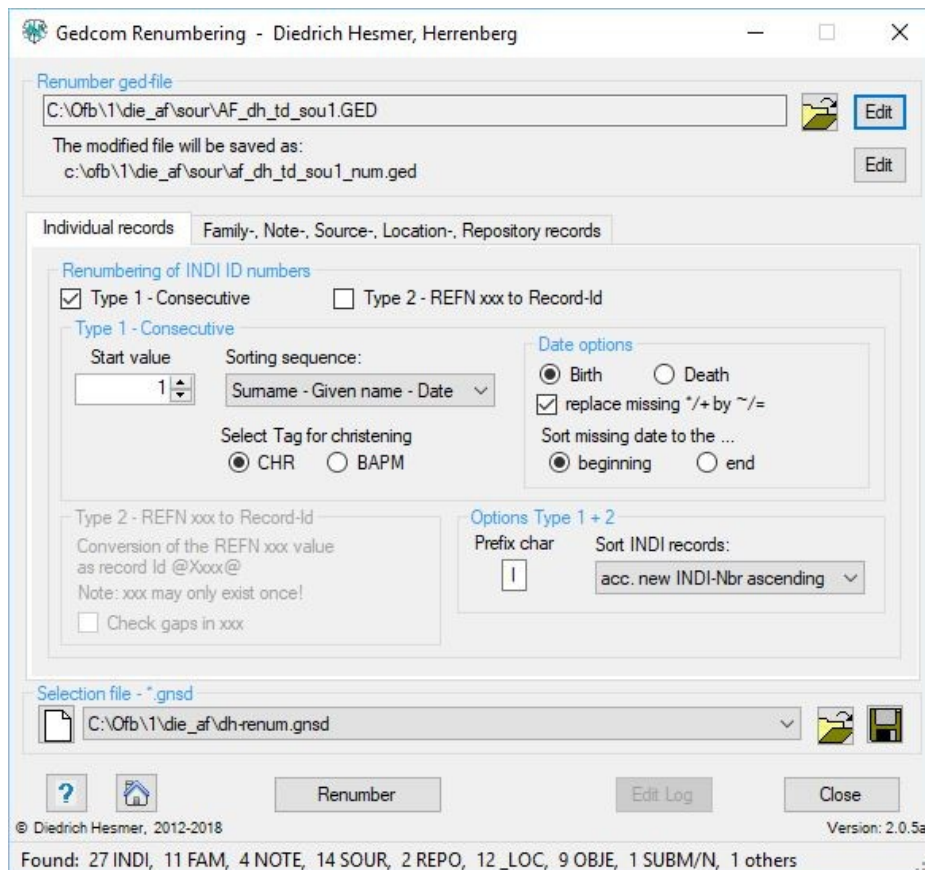

Fig 3.1: Main screen Gedcom Renumbering

● Pushing the respective 📂 a new ged-file or gnsd-file will be loaded.

● Pushing *"Edit"* by the left mouse button opens the ged-file by the text editor for viewing, defined by the "Base-Settings" of the OFB resp. Gedcom Start Center. Alternatively the standard text editor will be used. Pushing the right mouse button starts "*GedShow*" with the ged-file.

● With the optional "*H*" button to the right of the loaded ged file the HEAD record of the ged file is displayed in the browser.

● Pushing 💾 a folder (e.g. "C:\Data\Gedcom\files\") and name for the renumbering definition file must be specified and the file stored as "gnsd-file". This file contains all settings specified by all the tabs for reuse.

● Pushing 🗋 all fields will be set to their initial value.

● A click on the "DropDown" Element for the gnsd-selection file opens a list with the recent used gnsd files, up to the number defined at "Settings" in the Start-Center. Selecting one of these files will load this file.

● Pushing *?* opens the Gedcom Service Program user guide.

● Pushing the *house* opens the GSP homepage by the browser.

● Pushing *Renumber* the reading and renumbering of the source file and the writing of the renumbered ged-file starts. The modified file includes after the header all INDI records, then all FAM records followed by the NOTE, SOUR, REPO, OBJE, _LOC records in their original sequence. At the end all other none standard records will follow.
Note: Are orphaned references found during reading, they will be listed in a message window and a decision has to be made about a further processing. These are referenced record numbers, but the record number doesn't exist in the ged-file. For example, a family record with a "1 WIFE @I1234@" entry,

but there is no record "0 @I1234@ INDI". In case of a further processing a "_OR" will be appended for safety reasons. For example, a "1 WIFE @I1234@" will be modified to a "1 WIFE @I1234_OR@". This avoids that accidentally a valid reference will be produced by the renumbering process.

● Pushing **Edit log** opens the log-file by the text editor for viewing, defined by the "Base-Settings" of the OFB program resp. Gedcom Start Center. Alternatively the standard text editor will be used.

● Pushing **Close** the program exits and creates an ini-file containing the gnsd-filenames for reload.

● Messages will be written to the bottom of the screen.

3.3.1 Tab "Individual records"

Allows the selection and definitions for re-numbering and re-sorting of INDI records.

Two types of re-numbering are available:

● **Type 1 – Consecutive** performs a consecutive numbering of the records according to the following options:

  ○ **Initial value** defines the lowest value of the INDI number for renumbering to start from. Minimum is "0", standard is "1".

  ○ **Sorting sequence: Exclusively** the numbers of the INDI records for renumbering will be defined. The sequence of the records themselves will remain unchanged. Following sorting options are available:

     ■ no sorting
       the existing records will be renumbered ascending, starting with the initial value specified.

     ■ Surname – Given name – Date

     ■ Date – Surname – Given name

  ○ At **Date options** the method to find the sorting date will be specified.

     ■ **Birth** and **Death** are available

     ■ **replace missing birth/death date by christening/burial date** allows the replacement.

     ■ For **Sorting missing date** those may be set to the **beginning** (Date 00000000) or the **end** (Date 99999999) .

  ○ By **Select Tag for christening** the **"1 CHR"** and **"1 BAPM"** are available alternatively.

● **TYPE 2 – REFN xxx to Record-Id** converts the REFN xxx numbers found into record ID's. The xxx must be numeric, there must be one REFN tag in each record and no REFN xxx may occur multiple times in the file. This is checked and, if necessary, aborts the renumbering and listed in the log file.
  Additional options in addition to the *Options for type 1 + 2* listed below are:

  ○ **Check gaps in xxx** checks the found REFN numbers and lists missing numbers in the log file.

● **Options for type 1 + 2** are valid for both types:

  ○ Below of **Prefix char** enter the required capital letter (1 char !!). No letter may be used multiple times. This character will be added in front of the running number.

  ○ **Sort INDI records:** The sequence of the INDI records will be modified. Following sorting options are available:

     ■ unsorted – as original

     ■ acc. new INDI-Nbr ascending

3.3.2 Tab "Family-, Note-, Source-, Location-, Multimedia-, Repository records"

Fig 3.2: Tab "Family-, Note-, Source-, Location-, Repository records

Allows the selection and definition for re-numbering of FAM, NOTE, SOUR, REPO and _LOC records. Cryptic or unsorted record numbers may be modified to ascending numbers, where in addition 1 capital character will be added in front of the numbers.

Only such record types will be activated, which are present in the actual ged-file. Therefor during loading a ged-file will already be read and checked accordingly.

● Selection of *Family / Note / Location / Source / Multimedia / Repository records* activates the re-numbering of this type.

● Below of *Prefix char* enter the required capital letter (1 char !!). No letter may be added multiple times. This character will be added in front of the running number

● Below of *Start value* enter or select the starting value of the new numbering. Minimum is "0", standard is "1". From this value, the numbers are continuously incremented by 1.

By the *checking symbol* the lists can be selected/deselected.